EXHIBIT B

Page 1

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

```
_________________________________
IN RE MATTER OF:                 )
RICHARD KADREY, et al.,          )
Plaintiff,                       )
     vs.                         ) C.A. NO.:
META PLATFORMS, INC.,            ) 3:23-cv-03417-VC
Defendant.                       )
_________________________________)
```

VIDEOTAPED DEPOSITION OF YANN LeCUN, Ph.D.
Palo Alto, California
Thursday, November 21, 2024
** HIGHLY CONFIDENTIAL - ATTORNEYS EYES ONLY **
UNDER THE PROTECTIVE ORDER

Stenographically Reported by:
HEATHER J. BAUTISTA, CSR, CRR, RPR, CLR
Realtime Systems Administrator
California CSR License #11600
Oregon CSR License #21-0005
Washington License #21009491
Nevada CCR License #980
Texas CSR License #10725

DIGITAL EVIDENCE GROUP
1730 M. Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

Page 177

strategy? 14:31:17
A. I have no idea. 14:31:18
(Stenographer clarification.) 14:31:26
Q. (By Ms. Geman) Rather, do you know in what 14:31:26
form the -- the training data is used when -- when 14:31:29
portions of it are used for the memorization 14:31:34
limitation strategy? 14:31:38
A. It's the same format that is used during 14:31:40
training and I really don't know. I mean, my guess 14:31:43
is that it's just plain text for a lot of those 14:31:45
things. 14:31:48
Q. Are you familiar with the phenomenon where 14:31:54
LLaMA will, in response to a prompt, output 14:31:57
copyright-protected text and then it essentially 14:32:03
disappears? 14:32:06
A. Disappears? 14:32:08
Q. Goes off, like it no longer shows up on 14:32:08
the -- on the output. 14:32:11
MR. WEINSTEIN: Object to form. 14:32:14
Q. (By Ms. Geman) Sort of like Snapchat? 14:32:15
A. I don't understand the scenario. 14:32:19
Q. That somebody does a prompt and says, you 14:32:23

Page 178

know, give me a -- the first three paragraphs of X. 14:32:26

A. Uh-huh. 14:32:33

Q. And then LLaMA spits out the first three 14:32:33
paragraphs of X, but then it takes it back. 14:32:36

A. Oh, I see. 14:32:39

This -- this could occur, I believe, 14:32:42
because -- because there are systems that are put in 14:32:47
place that check whether the continuous output of 14:32:57
the LLM either is toxic or inappropriate or 14:33:04
infringes copyright and then takes down the text if 14:33:10
it does. And that's done, of course, after the text 14:33:14
is generated, but the possibility after the text has 14:33:17
been shown to the user for a short time. 14:33:27

Q. Do you know what that systems [sic] are 14:33:33
called? 14:33:35

A. No. 14:33:36

Q. And do you know if they're put in place 14:33:37
after validation of the model? 14:33:40

MR. WEINSTEIN: Object to form. 14:33:45

Q. (By Ms. Geman) Or let me rephrase the 14:33:46
question. 14:33:48

At what -- at what stage are those systems 14:33:48

Page 179

that check continuous output put in place? 14:33:51

MR. WEINSTEIN: Object to form. 14:33:57

THE WITNESS: Well, they're part of -- 14:33:58
okay. 14:34:00

So you have the LLM parts of the system; 14:34:00
right? So the system like Meta AI or any chatbot is 14:34:03
a complex system of multiple modules in interaction. 14:34:09
One of them is the LLM and you can use the LLM to 14:34:14
produce either a single text or multiple text using 14:34:18
the temperature adjustment. 14:34:22

You can also change the prompts so you can 14:34:24
add a hidden-system prompt to the user prompt to 14:34:27
bias the system towards generating certain types of 14:34:32
outputs. Okay. 14:34:36

So then the system generates multiple 14:34:36
outputs and then you can run those through systems 14:34:39
that are downstream that checks whether those 14:34:42
outputs are correct or incorrect, toxic or 14:34:44
non-toxic, you know, copyrighted material or not, 14:34:48
and it's -- it's possible that this check takes a 14:34:52
while and is expensive which is why you let the 14:34:58
system produce the text to the user and then after 14:35:01

Page 180

the copyright check, for example, has been done, you 14:35:04
take down and suppress the output 14:35:07

Q. (By Ms. Geman) So could that check only 14:35:12
kind of jump in if -- even after multiple pages of a 14:35:13
copyright-protected text has shown up? 14:35:18

A. No, it doesn't take that long. 14:35:20

Q. How long does it take? 14:35:22

A. I don't know, but -- but it's on the order 14:35:23
of seconds. 14:35:26

Q. Okay. 14:35:27
And what is -- what is the name of that -- 14:35:27
I mean, I asked generally the name of the system. 14:35:29
Do you know particularly the name of -- of the 14:35:32
system that checks once the output has been 14:35:36
released? 14:35:39

A. No, I don't know. 14:35:40

Q. And how -- how does that -- that system 14:35:40
determine whether the output is copyright protected? 14:35:46

MR. WEINSTEIN: Calls for speculation. 14:35:51

THE WITNESS: So checking copyright is 14:35:55
something that Meta has done on its services for a 14:35:56
very long time. Because people post copyrighted 14:36:02